

## AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior listings and versions of claims in the application:

### **Listing of Claims:**

1. (Currently Amended) A method for performing a supervised learning process in an artificial intelligence environment including optimizing a database of sample records for the training and testing of a prediction algorithm for predicting the presence or absence of a specified medical condition in a patient, the method A computer medium of sound or image recognition comprising the steps of:

one or more sensors or receivers responsive to signals;

a computer operatively coupled to the one or more sensors, the computer comprising a central processing unit;

one or more memories, at least one of the one or more memories storing a software program comprising the steps of:

defining a set of one or more plurality of distributions of the known database records onto respective training and testing subsets;

using the defined set of distributions to train and test training and testing a first generation set of one or more prediction algorithms using the plurality of distributions of the database records and assigning a fitness score to each, each of said prediction algorithms being associated with a certain first different distribution of said database records;

assigning a fitness score to each of the prediction algorithms;

feeding the set of prediction algorithms to an evolutionary algorithm which generates a set of one or more second generation prediction algorithms and assigns a fitness score to each;

continuing to feed each generational set of prediction algorithms to the evolutionary algorithm until a termination event occurs, wherein said termination event is at least one of:

a prediction algorithm generated with a fitness score equal to or exceeding a defined minimum value,

the maximum fitness score of successive generational sets of prediction algorithms converging to a given value, and/or

a certain number of generations having been generated;

selecting a prediction algorithm having a best fitness score; and

using the distribution of database records associated with said selected prediction algorithm in performing supervised learning, said supervised learning including training and testing of prediction algorithms to obtain a trained prediction algorithm; wherein

~~said method is performed using a computer and computer software forming an intelligent system, and~~

~~the trained prediction algorithm is effective to predict output variables for data relating to said condition, thereby predicting diagnosis of said condition,~~

~~and further comprising the steps of:~~

generating a population of prediction algorithms, wherein each ~~one~~ of said prediction algorithms is trained and tested according to a second different distribution of the records of the data set in the complete database onto a training data set and a testing data set,

~~each second different distribution being created as one of a random or pseudorandom distribution and a distribution formed by a deterministic mathematical process characterized as a pseudorandom distribution,~~

each prediction algorithm of said population being trained according to its own distribution of records of the training set and being validated in a blind way according its own distribution on the testing set, and

a score reached by each prediction algorithm being calculated in the testing phase representing its fitness;

providing an evolutionary algorithm which combines the different models of distribution of the records of the complete data set in a training and in a testing set, which sets are represented each one by a corresponding prediction algorithm trained and tested on the basis of said training and testing data set according to the fitness score calculated in the previous step for the corresponding prediction algorithm,

the fitness score of each prediction algorithm corresponding to one of the different distributions of the complete data set on the training and the testing data sets being the probability of evolution of each prediction algorithm or of each said distribution of the complete data set on the training and testing data sets;

repeating the evolution of the prediction algorithm generation for a finite number of generations or till the output of the genetic algorithm converges to a best solution and/or till the fitness value of at least some prediction algorithm related to an associated data records distribution has reached a desired value; and

setting the data records distribution for the best solution as the optimized training and testing subsets for training and testing prediction algorithm; and

an output system providing an indication of the signals detected by the one or more sensors.

2. (Canceled)

3. (Currently Amended) ~~A method~~-The computer medium according to claim 1, wherein to each record of the data set ~~wherein the software program further comprises the step of associating~~ a distribution variable is ~~associated to~~ each record of the data set, which is binary and ~~which~~ has at least two ~~status statuses~~, one of ~~this the~~ two ~~status statuses~~ being associated with the inclusion of the record in the training set and the other ~~one of the two statuses~~ in the testing set.

4. (Previously Presented) ~~A method~~-The computer medium according to claim 1, wherein the prediction algorithm is an artificial neural network.

5. (Previously Presented) The computer medium~~A method~~ according to claim 1, wherein the prediction algorithm is a classification algorithm.

6. (Currently Amended) ~~A method~~The computer medium according to claim 1, wherein once an optimum distribution has been computed, the optimized training data subset is made equal to a complete data set, ~~being~~ the individuals included in the training subset ~~being~~ distributed onto a new training set and onto a new testing set each one having about the half of the records of the original optimized training set, while the originally optimized testing set is used as a third data subset for validation purposes.

7. (Currently Amended) ~~A method~~The computer medium according to claim 6, wherein the distribution of the data of the originally optimized training set onto the new training and new testing set is optimized ~~by means of~~through a pre-processing phase including the steps of said method for optimizing a database of sample records, said records being records in the originally optimized training set.

8. (Currently Amended) ~~A method~~The computer medium according to claim 1, wherein different choices of the structure of the training subset and the structure of the testing subset ~~consist in~~comprise different selections of the number of input variables of the data records of the database, which selections ~~consist in~~include leaving out at least one, ~~preferably~~ two or more variables variable from the entire input variable set forming each record, the records of the database comprising a certain number of known input variables and a certain number of known output variables.

9. (Currently Amended) ~~A method~~The computer medium according to claim 8, further ~~comprising~~comprising the following steps:

defining a distribution of data from the complete data set onto a training data set and onto a testing data set;

generating a population of different prediction algorithms each one having a training and/or testing data set in which only some variables have been considered among all the original variables provided in the data sets, each one of the prediction algorithms being

generated by means of through a different selection of variables;

carrying out learning and testing of each prediction algorithm of the population and evaluating the fitness score of each prediction algorithm;

applying an evolutionary algorithm to the population of prediction algorithms for achieving new generations of prediction algorithmalgorithms;

for each generation of new prediction algorithms, representing each one a new different selection of input variablevariables, testing or validating the best prediction algorithm according to the best hypothesis of input variables selection; and

evaluating a fitness score and promoting the prediction algorithms, representing the selections of input variables which have the best testing performances and the minimum input variables, for the processing of the new generations.

10. (Previously Presented) ~~A method~~ The computer medium according to claim 8, further comprising a preprocessing phase, including the steps of said method for optimizing a database of sample records, for selecting the most predictive input variables.

11. (Currently Amended) ~~A method~~ The computer medium according to claim 1,

in which different choices of the structure of the training subset and the structure of the testing subset ~~consist in comprise~~ different selections of the number of input variables of the data records of the database, which selections ~~consist in include~~ leaving out at least one, preferably two or more variables variable from the entire input variable set forming each record, the records of the database comprising a certain number of known input variables and a certain number of known output variables,

and further comprising a pre-processing phase, including the steps of said method for optimizing a database of sample records, for selecting the most predictive input variables,

wherein the database subjected to the [[a]] pre-processing phase of input variable selection is a training subset and a testing subset processed with said method.

12. (Currently amended) ~~A method~~ The computer medium according to claim 1, wherein the complete database the distribution of the records of which has to be optimized has

data records having a selected number of input variables, the selection being carried out with said method, and in which-wherein different choices of the structure of the training subset and the structure of the testing subset consist in comprise different selections of the number of input variables of the data records of the database, which selections consist in leaving out at least one, preferably two or more variables variable from the entire input variable set forming each record, the records of the database comprising a certain number of known input variables and a certain number of known output variables.

13. (Previously Presented) A-method-The computer medium according to claim 1, wherein a preprocessing phase for optimizing the distribution of the records on a training subset and a testing subset and for selecting the most predictive input variables, is carried out alternatively one to the other several times.

14. (Currently Amended) A-method-The computer medium according to claim 1, wherein the evolutionary algorithm is a genetic algorithm with the following evolutionary rules:

an average health value of the population is computed as a function of the fitness values of each single individual in the population;

coupling, recombination of genes and mutation of genes are carried out in a differentiated manner depending on a comparison between the fitness of each individual of the couple and the average health value of the entire population to which the individuals belong;

individuals having a fitness value lower or equal to the average health of the entire population are not excluded from the creation of new generations but are marked out and entered in a vulnerability list; and

the number of subjects entered in the vulnerability list defining-defines the number of possible marriages.

15. (Currently Amended) A-method-The computer medium according to claim 14, wherein-in which for coupling purposes and for generation of children at least one parent individuals must have a fitness value greater than the average health value of the population.

16. (Currently Amended) ~~A method~~-The computer medium according to claim 14, wherein each couple of individuals ~~can~~are adapted to generate ~~individuals offsprings~~ having a fitness different from the average health, ~~so called offsprings~~ if the fitness of one them[[.]] at least is greater than the average fitness, the offsprings of each marriage occupying the places of subjects entered in the vulnerability list and ~~are~~ marked out, so that a weak individual can continue to exist through his own children.

17. (Previously Presented) ~~A method~~-The computer medium according to claim 14, wherein coupling between individuals having a very low fitness value and a very high fitness value are not allowed.

18. (Currently Amended) ~~A method~~-The computer medium according to claim 14, wherein the following recombination rules of the genes of the ~~coupled parent~~parents individuals ~~coupled~~ are considered in the case the parents individuals have ~~not no~~ common genes:

the health of father and mother individuals are greater than the average health of the entire population;

the crossover is a classical crossover according to which the genes of the father and of the mother individuals are substituted one with the other starting from a certain crossover point;

the health of father and mother individuals are lower than the average health of the entire population[[.]], in this case the two children are formed through rejection of the parents' genes they will receive by the crossover process;

the health of one of the parents is less than the average health of the entire population while the health of the other parent is greater than the average health of the entire population[[.]], in this case only the parents whose health is greater than the average health of the entire population will transmit their genes, while the genes of the parent having an health lower than the average health of the entire population are rejected.

19. (Currently Amended) ~~A method~~-The computer medium according to claim 18, wherein each gene is ~~characterised characterized~~ by a status level, ~~the method further and~~ wherein ~~genes gene~~ rejection ~~consists in comprises~~ modifying the status of the genes from one status level to a different status level.

20. (Currently Amended) ~~A method~~ The computer medium according to claim 18, wherein a modified crossover of the genes of the ~~parents~~ parent individuals is carried out when the ~~parents~~ parent individuals ~~has~~ have part of the genes that coincide, this modified crossover provides providing for generating and an offspring in which the genes selected for crossover are the most effective ones of the parents.

21. (Currently Amended) ~~A method~~ The computer medium according to claim 14, wherein the individuals are the different prediction algorithm algorithms representing a corresponding different initial random distribution of data records onto the testing and the training data set, and wherein the genes consist in the binary status variable of association of each record to the training and to the testing subset.

22. (Currently Amended) ~~A method~~ The computer medium according to claim 14, wherein the individuals are the prediction algorithms each one representing a different training and testing data set, the difference residing in a different selection of input variables for each different training and testing subset, and wherein the genes consist in the comprise a different selection variable which is provided for each input variable in the different training and testing subsets, the above-mentioned selection variable being a parameter indicating the presence/absence of each corresponding input variable in the records of each data set.

23. - 26. (Canceled)

27. (Currently Amended) ~~The system~~ computer medium according to claim [[25]] 1, wherein the ~~system~~ is a system for image recognition, the input means being responsible to electromagnetic waves, the system being able to output is an indication of a recognize the shape of an object generating or reflecting electromagnetic waves, and/or the distance and/or the identity of the object.

28. (Currently Amended) ~~The system~~ computer medium according to claim [[26]] 1,

wherein the database of known database records comprises comprise acoustic signals emitted by one or more objects or one or more living beings making part of the-a typical environment in which the device has to operate or the method is performed or data relating to one or more images of one or more objects or one or more living beings making that are part of the typical environment, in which the device has to operate to which are univoquely correlated to corresponding known kind, and/or identity and/or meaning of objects to which the said acoustic signals or image data are related and/or from which the-said acoustic signals or image data are generated.

29. (Currently Amended) The system computer medium according to claim 27, wherein the system computer medium is a specialized system for image pattern recognition having artificial intelligence utilities for analyzing a-digitalized image, i.e. an image in the form of a array of image data records, each image data record being related to a zone or point or unitary area or volume of a two or three dimensional visual image, the visual image being formed by an array of the said pixels or voxels and utilities for indicating for each image data record a certain quality among a plurality of known qualities of the image data records;

the system having a processing unit, a memory in which an image pattern recognition algorithm is stored in the form of a software program which can be executed by the processing unit;

a memory in which a certain number of predetermined different qualities which the image data records can assume has been stored and which qualities has to be univoquely associated to each of the image data records of an image data array fed to the system;

input means for receiving wherein the one or more sensors or receivers receive arrays of digital image data records or input means for generating arrays generate an array of digital image data records from an existing image;

wherein at least one of one of the one or more memories stores and a memory for storing the said digital image data array.[:] and

output means for indicating wherein the output system indicates for each image data record of the image data array a certain quality chosen by the processing unit in carrying out the image pattern recognition algorithm in the form of the said software program.[:]

the image pattern recognition algorithm is a prediction algorithm in the form of a software

program, which prediction algorithm is further associated to a system being further provided with a training and testing software program;

the system is able to carry out training and testing according to the method of claim 1;

the method is provided in the system in the form of the training and testing software program; and

a database being also provided in which data records are contained univocally associating known image data records of known image data arrays with the corresponding known quality from a certain number of predetermined different qualities which the image data records can assume.

30. (Canceled)

31. (Currently Amended) A method-computer medium according to claim [[30]] 38, wherein an optimization of the distribution of the records of the original database in a training dataset and in a testing dataset is carried out in one of a pre processing and a post processing phase, i.e. before carrying out the steps e) to m) at step d) or after having carried out the steps a) to m).

32. – 35. (Canceled)

36. (New) The computer medium according to claim 1, wherein the signals are electromagnetic waves in the acoustic or visible range.

37. (New) The computer medium according to claim 1, wherein the software program further comprises a preprocessing phase comprising the steps of:

defining a plurality of distributions of the records of the optimized training subset onto new training and testing subsets;

training and testing a new generation set of prediction algorithms using the new training and testing subsets;

assigning a fitness score to each prediction algorithm in the new generation of prediction algorithms;

defining a new optimized training subset and a new optimized testing subset;

identifying a new optimized training subset and a new optimized testing subset as the training and testing subsets corresponding to the prediction algorithm having the highest fitness score; and

employing the optimized testing subset as a validation set.

38. (New) A computer medium for producing a microarray for genotyping, the computer medium comprising:

a computer comprising a central processing unit;

one or more memories, at least one of the one or more memories storing a database of experimentally determined data in which each record relates to a known clinical or experimental case of a sample population of cases, the data comprising a number of input variables corresponding to the presence/absence of a predetermined number of polymorphisms and/or mutations and/or equivalent genes of a number of theoretically probable relevant genes, said certain predetermined number of polymorphisms and/or genes forming a set, and the data further comprising one or more related output variables corresponding to the certain biological or pathologic condition of the clinical and experimental cases of the sample population;

at least one of the one or more memories storing a software program defining a number of theoretically relevant genes or alleles or polymorphisms relevant for a biologic condition, the software program comprising the steps of:

determining a selection of a subset of the set of certain predetermined number of polymorphisms and/or genes by testing the association of the genes or polymorphisms and the biological or pathological condition by mathematical tools comprising a prediction algorithm applied to the database;

defining a plurality of distributions of the database onto respective training and testing subsets;

training and testing a first generation set of prediction algorithms using the plurality of distributions of the database, each of said prediction algorithms being

associated with a first different distribution of records of the database;

assigning a fitness score to each of the prediction algorithms;

feeding the set of prediction algorithms to an evolutionary algorithm which generates a set of one or more second generation prediction algorithms and assigns a fitness score to each;

continuing to feed each generational set of prediction algorithms to the evolutionary algorithm until a termination event occurs, wherein said termination event is at least one of:

- a prediction algorithm generated with a fitness score equal to or exceeding a defined minimum value,
- the maximum fitness score of successive generational sets of prediction algorithms converging to a given value, or
- a certain number of generations having been generated;
- selecting a prediction algorithm having a best fitness score; and
- using the distribution of database associated with said selected prediction algorithm in performing supervised learning, said supervised learning including training and testing of prediction algorithms to obtain a trained prediction algorithm;

generating a population of prediction algorithms, wherein each of said prediction algorithms is trained and tested according to a second different distribution of the records of the data set in the complete database onto a training data set and a testing data set,

each second different distribution being created as one of a random or pseudorandom distribution,

each prediction algorithm of said population being trained according to its own distribution of records of the training set and being validated in a blind way according its own distribution on the testing set, and

a score reached by each prediction algorithm being calculated in the testing phase representing its fitness;

providing an evolutionary algorithm which combines the different models of distribution of the records of the complete data set in a training and in a testing set,

which sets are represented each one by a corresponding prediction algorithm trained and tested on the basis of said training and testing data set according to the fitness score calculated in the previous step for the corresponding prediction algorithm,

the fitness score of each prediction algorithm corresponding to one of the different distributions of the complete data set on the training and the testing data sets being the probability of evolution of each prediction algorithm or of each said distribution of the complete data set on the training and testing data sets;

repeating the evolution of the prediction algorithm generation for a finite number of generations or till the output of the genetic algorithm converges to a best solution and/or till the fitness value of at least some prediction algorithm related to an associated data records distribution has reached a desired value; and

setting the data records distribution for the best solution as the optimized training and testing subsets for training and testing prediction algorithm; and  
an output system responsive to the received information.